|  |  |
| --- | --- |
| 1机器学习算法：学习10十大常用算法原理和思想？(1.决策树2.随机森林算法3.逻辑回归4.SVM5.朴素贝叶斯6.K最近邻算法7.K均值算法8.Adaboost 算法9.神经网络10马尔可夫)  学习那一种或几种常用的人工智能算法库;  tensorflow,weka,matlab，hadoop的mahout，还是阿里的自己的机器学习算法库。  对机器学习算法的要求等。 | 由于机器学习内容相对来说多，而且目前分支也比比较多。因为机器学习这个已经研究了很多年了，分支也比较多，理论性的多，学习起来还是有些枯燥和繁琐。  如果能提供一些路线会非常好。 |
| 2.主要用什么语言做开发python，java，R语言（我的java熟悉些）。对用R语言处理大数据要求高不高，还有数据挖掘部分有要求没有呢？ |  |
| 3.对机器学习的一些指导方法，推荐一些书，推荐一些博客或网站。 |  |
| 4.工作是开发还是研究算法运用。 |  |
| 5.提供一些实验数据源，和实验平台 |  |

一：机器学习算法

|  |  |
| --- | --- |
| 我对大数据的理解是：（大致分为三个方向） | |
| 1，大数据开发工程师，偏重编码，sql语句，mapreduce的开发。spark函数式编程   1. 数据挖掘工程师和数据分析师：偏重统计学和数据挖掘技术   3.人工智能算法工程师，学习研究机器学习算法。参数调优等。 | 左边这三个方向应该向那个方向努力。  1.大数据开发工程师，是我之前自学过一段时间的。  2.还没有怎么接触  3.自学过一点神经网络，卷积神经网络。 |
| 之前学习的javaEE，用来做网站，感觉那个以后做的人太多，而且门槛低，没有太大竞争优势，所以自己在学习大数据相关的知识。  由于不了解现在企业的需求，所以之前把大部分时间和精力花在hadoop集群搭建，hadoop的hdfs，yarn，mapreduce，zookeeper，spark等技术和知识，前段时间已经完成在hadoop环境上跑了一些基本实验，现在有些迷茫。在考虑是否有必要深入去hadoop底层知识，还是有strom流式计算等技术。  因为精力和时间有限，深知不能广撒网，而且理论学习太多，有些打消积极性。所以需要一些方向性的指导。 |  |
| * 正则化算法（Regularization Algorithms） * 集成算法（Ensemble Algorithms） * 决策树算法（Decision Tree Algorithm） * 回归（Regression） * 人工神经网络（Artificial Neural Network） * 深度学习（Deep Learning） * 支持向量机（Support Vector Machine） * 降维算法（Dimensionality Reduction Algorithms） * 聚类算法（Clustering Algorithms） * 基于实例的算法（Instance-based Algorithms） * 贝叶斯算法（Bayesian Algorithms） * 关联规则学习算法（Association Rule Learning Algorithms） * 图模型（Graphical Models） * 岭回归（Ridge Regression） * 最小绝对收缩与选择算子（LASSO） * GLASSO * 弹性网络（Elastic Net） * 最小角回归（Least-Angle Regression） | 这些算法系统学起来，如果深入，一两年都不够，怎样去学习可以达到企业的要求。完全小白，所以问题有点多。 |

（需要提供一些学习方向性的指导。早做准备，明确目标，然后购买书籍。）

二、关于编程能力

关于基础编程能力，这个应该还可以。以前有做过ACM竞赛。再花点时间复习，应该没有大问题。